OpenCAPI and its Roadmap

Myron Slota, President
OpenCAPI Consortium

OpenPOWER Summit Europe
RAI Centre | Amsterdam
October 3-4, 2018

Join the Conversation #OpenPOWERSummit
Industry Collaboration and Innovation

OpenCAPI

A data-centric approach to server design
OpenCAPI and its Roadmap

Topics

- Industry Background and its influence on IBM’s IO and microprocessor roadmap
- Limitations of CAPI running over PCIe
- Creation of the OpenCAPI architecture and its benefits
- OpenCAPI Consortium
- Potential solutions based on coherent attached accelerators
Industry Background that Defined OpenCAPI

- Emerging workloads driving computational demand (e.g., AI, cognitive computing)
- Moore’s Law not being supported by traditional silicon scaling
- Driving increased dependence on Hardware Acceleration for performance to continue on Moore’s Law
  - **Hyperscale Datacenters and HPC**
    - Require much higher network bandwidth (100 Gb/s -> 200 Gb/s -> 400Gb/s are emerging)
  - **HPC and Artificial Intelligence** *(deep learning, inferencing)*
    - Need more bandwidth between accelerators and memory
  - **Emerging memory/storage technologies**
    - Driving need for bandwidth with low latency
Industry Background that Defined OpenCAPI

- Hardware accelerators are defining the attributes required of a high performance bus
  - Offload the computational demand on the microprocessor – if you are going to use accelerators, the bus needs to handle *large amounts of data quickly*
  - Growing demand for *network performance and offload*
  - Introduction of device *coherency requirements*
  - Emergence of *complex storage and memory solutions*
  - Supports all various form factors (e.g., GPUs, FPGAs, ASICs, etc.)

...all relevant to modern data centers
IBM’s IO Offering Legacy

- IBM IO offerings has always been leading edge (IBM's ‘PowerAccel’ family)

- Of most recent
  - PCIE Advanced Generations (first in Gen4)
  - NVLink (first of its kind proprietary link to NVIDIA GPUs)
  - CAPI (coherent interface that ran over PCIe)
  - OpenCAPI (a new architecture)
POWER Processor Technology Roadmap

POWER7 45 nm

- Enterprise
  - 8 Cores
  - SMT4
  - eDRAM L3 Cache

POWER7+ 32 nm

- Enterprise & Big Data Optimized
  - Up to 12 Cores
  - SMT8
  - CAPI Acceleration
  - High Bandwidth GPU Attach

POWER8 Family 22 nm

- POWER9 Family 14 nm

Built for the Cognitive Era
- Enhanced Core and Chip Architecture Optimized for Emerging Workloads
- Processor Family with Scale-Up and Scale-Out Optimized Silicon
- Premier Platform for Accelerated Computing

1H10

2H12

1H14 – 2H16

2H17 – 2H18+
OpenCAPI and its Roadmap

POWER8

- Felt a need to introduce to industry of a coherent interface for accelerators
- The IO offering expanded
  - Higher performance with PCIe Gen3
  - Introduced CAPI 1.0 (Coherent Accelerator Processor Interface) that runs over PCIe
- CAPI 1.0 Acceptance
  - Power Architecture specific
  - Improvement needed in ease of implementation (i.e., PSL), use and programming
  - Newness and limited products that leveraged this technology
  - However, proved there was Industry interest to drive accelerators
OpenCAPI and its Roadmap

POWER9

- Continued expansion of the IO offering
- Higher performance with **PCIe Gen4** and **NVLink 2.0**
- Given the validation for the need of a coherent high performance bus

  - IBM worked to improve the CAPI architecture that runs over PCIe and introduced **CAPI 2.0**
    - Improved bandwidth performance by ~3X
    - Reduced overhead complexity in the accelerator design address translation on host resulting in more efficiency

  - In addition, IBM introduced a new IO architecture and released the **OpenCAPI 3.0 and OpenCAPI 3.1** specifications
Why OpenCAPI and what is it?

- OpenCAPI is a new ‘bottom’s up’ IO standard

- Key Attributes of OpenCAPI 3.0
  - **Open IO Standard** – *Choice for developers and others to contribute and grow an ecosystem*
  - **Coherent interface** – *Microprocessor memory, accelerator and caches share the same memory space*
  - **Architecture agnostic** – *Capable going beyond Power Architecture*
  - **High performance** – *No OS/Hypervisor/FW Overhead for Low Latency and High Bandwidth*
  - **Not tied to Power** – *Architecture Agnostic*
  - **Ease of programming**
  - **Ease of implementation with minimal accelerator design overhead**
  - **Ideal for accelerated computing and SCM including various form factors (FPGA, GPU, ASIC, TPU, etc.)**
  - **Optimized for within a single system node**
  - **Supports heterogeneous environment – Use Cases**

- **OpenCAPI 3.1**
  - Applies OpenCAPI technology for use of standard DRAM off the microprocessor
  - Based on an Open Memory Interface (OMI)
  - Further tuned for extreme lower latency
Use Cases - A True Heterogeneous Architecture Built Upon OpenCAPI

1. **Accelerators**: The performance, virtual addressing and coherence capabilities allow FPGA and ASIC accelerators to behave as if they were integrated into a custom microprocessor.

2. **Coherent Network Controller**: OpenCAPI provides the bandwidth that will be needed to support rapidly increasing network speeds. Network controllers based on virtual addressing can eliminate software overhead without the programming complexity usually associated with user-level networking protocols.

3. **Advanced Memory**: OpenCAPI allows system designers to take full advantage of emerging memory technologies to change the economics of the datacenter.

4. **Coherent Storage Controller**: OpenCAPI allows storage controllers to bypass kernel software overhead, enabling extreme IOPS performance without wasting valuable CPU cycles.

OpenCAPI specifications are downloadable from the website at www.opencapi.org
- Register
- Download

OpenCAPI 4.0 now added to the web site!
## Proposed POWER Processor Technology and I/O Roadmap

<table>
<thead>
<tr>
<th>Year</th>
<th>Architecture</th>
<th>Cores</th>
<th>Technology</th>
<th>Micro-Architecture</th>
<th>Process Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>POWER7</td>
<td>8</td>
<td>45nm</td>
<td>New Micro-Architecture</td>
<td>New Process Technology</td>
</tr>
<tr>
<td>2012</td>
<td>POWER7+</td>
<td>8</td>
<td>32nm</td>
<td>Enhanced Micro-Architecture</td>
<td>New Process Technology</td>
</tr>
<tr>
<td>2014</td>
<td>POWER8</td>
<td>12</td>
<td>22nm</td>
<td>New Micro-Architecture</td>
<td>New Process Technology</td>
</tr>
<tr>
<td>2016</td>
<td>POWER8 w/ NVLink</td>
<td>12/24</td>
<td>14nm</td>
<td>Enhanced Micro-Architecture</td>
<td>With NVLink</td>
</tr>
<tr>
<td>2017</td>
<td>P9 SO</td>
<td>12/24</td>
<td>14nm</td>
<td>New Micro-Architecture</td>
<td>Direct attach memory</td>
</tr>
<tr>
<td>2018</td>
<td>P9 SU</td>
<td>12/24</td>
<td>14nm</td>
<td>Enhanced Micro-Architecture</td>
<td>Buffered Memory</td>
</tr>
<tr>
<td>2019</td>
<td>P9 w/ Adv. I/O</td>
<td>12/24</td>
<td>14nm</td>
<td>Enhanced Micro-Architecture</td>
<td>New Memory Subsystem</td>
</tr>
<tr>
<td>2020+</td>
<td>P10</td>
<td>TBA</td>
<td>N/A</td>
<td>New Micro-Architecture</td>
<td>New Technology</td>
</tr>
</tbody>
</table>

### Sustained Memory Bandwidth
- **POWER7**: 65 GB/s
- **POWER8**: 210 GB/s
- **POWER9**: 150 GB/s
- **POWER10**: 350+ GB/s

### Standard I/O Interconnect
- **POWER7**: PCIe Gen2
- **POWER8**: PCIe Gen3
- **POWER9**: PCIe Gen4 x48
- **POWER10**: PCIe Gen5

### Advanced I/O Signaling
- **POWER7**: N/A
- **POWER8**: 20 GT/s 160GB/s
- **POWER9**: 25 GT/s 300GB/s
- **POWER10**: 25 GT/s 300GB/s

### Advanced I/O Architecture
- **POWER7**: N/A
- **POWER8**: CAPI 1.0, NVLink
- **POWER9**: CAPI 2.0, OpenCAPI3.0, NVLink
- **POWER10**: CAPI 2.0, OpenCAPI4.0, NVLink

---

**Statement of Direction, Subject to Change**
Development Activities Leading to Solutions with Development Partners

- Storage Class Memory
- DRAM Memory Solutions
- FPGA Accelerators
- GPUs
- Smart NICs
- ASICs
- Systems
Comparison of Memory Paradigms

- Common physical interface between non-memory and memory devices
- OpenCAPI protocol was architected to minimize latency; excellent for classic DRAM memory
- Extreme bandwidth beyond classical DDR memory interface
- Agnostic interface will handle evolving memory technologies in the future (e.g., compute-in-mem)
- Ability to handle a memory buffer to decouple raw memory and host interface to optimize power, cost, perf

**Main Memory**

Example: Basic DDR attach

**Emerging Storage Class Memory**

Storage Class Memories have the potential to be the next disruptive technology..... Examples include ReRAM, MRAM, Z-NAND...... All are racing to become the defacto

**Tiered Memory**

Storage Class Memory tiered with traditional DDR Memory all built upon OpenCAPI 3.1 & 3.0 architecture. Still have the ability to use Load/Store Semantics

OpenCAPI 3.1 Architecture
Ultra Low Latency ASIC buffer chip adding +5ns on top of native DDR direct connect!!
Acceleration Paradigms with Great Performance

Examples: Machine or Deep Learning such as Natural Language processing, sentiment analysis or other Actionable Intelligence using OpenCAPI attached memory

Examples: Encryption, Compression, Erasure prior to delivering data to the network or storage

Examples: Database searches, joins, intersections, merges. Only the Needles are sent to the processor

OpenCAPI is ideal for acceleration due to Bandwidth to/from accelerators, best of breed latency, and flexibility of an Open architecture

Examples: NoSQL such as Neo4J with Graph Node Traversals, etc

Examples: Video Analytics, Network Security, Deep Packet Inspection, Data Plane Accelerator, Video Encoding (H.265), High Frequency Trading etc
OpenCAPI Consortium
• Open forum founded by AMD, Google, IBM, Mellanox, and Micron
  • Manage the OpenCAPI specification, Establish enablement, Grow the ecosystem
  • Currently over 35 members
  • Why its own consortium? Architecture agnostic thus capable of going beyond Power Architecture

• Consortium now established
  • Established Board of Directors (AMD, Google, IBM, Mellanox Technologies, Micron, NVIDIA, Western Digital, Xilinx)
  • Governing Documents (Bylaws, IPR Policy, Membership) with established Membership Levels
  • Technical Steering Committee with Work Group Process established
  • Marketing/Communications Committee
  • Website www.opencapi.org
OpenCAPI Consortium

• **Work Groups**
  - TL Specification, DL Specification, PHY Signaling, PHY Mechanical, Compliance, and Enablement
  - Creation of additional work groups include: Memory/Accelerator, Software, and more
  - Design enablement available today (reference designs, documentation, SIM environment, exercisers, etc.)
  - OpenCAPI 3.0 and 3.1 Specifications available on web site was contributed to consortium as starting point for the Work Groups

• **OpenCAPI 4.0 now added to the web site !**
  - AFU Coherent Data Caching of System Memory
  - AFU Address Translation Caching (allows posted operations to system memory)
Cross Industry Collaboration and Innovation

Welcoming new members in all areas of the ecosystem
<table>
<thead>
<tr>
<th>Level</th>
<th>Cost</th>
<th>Entitlement Details</th>
</tr>
</thead>
</table>
| Strategic level  | $25K   | • Draft and Final Specifications and enablement  
• License for Product development  
• Workgroup participation and voting  
• TSC participation  
• Vote on new Board Members  
• Nominate and/or run for officer election  
• Prominent listing in appropriate materials |
| Contributor level| $15K   | • Draft and Final Specifications and enablement  
• License for Product development  
• Workgroup participation and voting  
• TSC participation  
• Submit proposals |
| Observing level  | $5K    | • Final Specifications and enablement  
• License for Product development |
| Academic and Non-Profit level | Free   | • Final Specifications and enablement  
• Workgroup participation and voting |
## Table of Enablement Deliveries

<table>
<thead>
<tr>
<th>Item</th>
<th>Delivery Name</th>
<th>Where to Obtain</th>
<th>Available When</th>
</tr>
</thead>
<tbody>
<tr>
<td>OpenCAPI 3.0 TLx and DLx Reference Xilinx FPGA Designs (RTL and Specifications)</td>
<td>&lt;snapshot&gt;.tar.gz</td>
<td>Enablement WG</td>
<td>Today</td>
</tr>
<tr>
<td>Xilinx Vivado Project Build with Memcopy Exerciser</td>
<td>Vivado Project Flow</td>
<td>Enablement WG</td>
<td>Today</td>
</tr>
<tr>
<td>Device Discovery and Configuration Specification and RTL</td>
<td>OpenCAPI 3.0 Configuration Sub-System Reference Design Specification</td>
<td>Enablement WG Causeway</td>
<td>Today</td>
</tr>
<tr>
<td>AFU Interface Specification</td>
<td>TLX 3.0 Reference Design.pdf</td>
<td>Enablement WG Causeway</td>
<td>Today</td>
</tr>
<tr>
<td>25Gbps PHY Mechanical Specification</td>
<td>25Gbps Interface Mechanical Spec</td>
<td>PHY Mechanical WG Causeway</td>
<td>Today</td>
</tr>
<tr>
<td>OpenCAPI Simulation Environment (OCSE)</td>
<td>ocse-&lt;version&gt;.tar.gz, OpenCAPIDemokit.pdf</td>
<td>Enablement WG</td>
<td>Today</td>
</tr>
<tr>
<td>Memcopy and Memory Home Agent Exercisers</td>
<td>MCP3 and LPC &lt;snapshot&gt;.tar.gz</td>
<td>Enablement WG</td>
<td>Today</td>
</tr>
<tr>
<td>Reference Driver Available</td>
<td>LIBOCXL</td>
<td>Ubuntu 18.04 GitHub</td>
<td>Today</td>
</tr>
</tbody>
</table>
## OpenCAPI Committee and Work Group Status

<table>
<thead>
<tr>
<th>Item</th>
<th>Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>OpenCAPI Technical Steering Committee</td>
<td>Up and running</td>
</tr>
<tr>
<td>Marketing &amp; Communications Committee</td>
<td>Up and running</td>
</tr>
<tr>
<td>PHY Signaling Workgroup</td>
<td>Up and running</td>
</tr>
<tr>
<td>PHY Mechanical Workgroup</td>
<td>Up and running</td>
</tr>
<tr>
<td>TL Architecture Specification Workgroup</td>
<td>Up and running</td>
</tr>
<tr>
<td>DL Architecture Specification Workgroup</td>
<td>Up and running</td>
</tr>
<tr>
<td>Enablement Workgroup</td>
<td>Up and running</td>
</tr>
<tr>
<td>Compliance Workgroup</td>
<td>Up and running</td>
</tr>
<tr>
<td>Accelerator Workgroup</td>
<td>Forthcoming</td>
</tr>
<tr>
<td>Memory Workgroup</td>
<td>Forthcoming</td>
</tr>
</tbody>
</table>
SmartDV

OpenCAPI Verification IP

Benefits
- Complete Verification of OpenCAPI Design
- Easy to Use
- Simplify Result Analysis
- Runs in every sim environment

SmartDV OpenCAPI VIP Environment contains:
- Complete regression suite
- Usage examples
- Detailed documentation
- User’s Guide and Release Notes

JOIN TODAY!

www.opencapi.org

COME SEE US AT THE OPENCAPI BOOTH
Thank You!
Questions?